DOCUMENT RESUME

ED 301 599 TM 012 592

AUTHOR Lum, John B.
TITLE K-12 Standardized Achievement and Aptitude Testing: The End Users.
PUB DATE Apr 87
NOTE 13p.; Testimony presented at the Annual Conference of the National Association of Asian and Pacific American Education (9th, Honolulu, HI, April 11, 1987).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Aptitude Tests; Asian Americans; *Educational Policy; Elementary School Students; Elementary Secondary Education; Longitudinal Studies; Pacific Americans; Parents; Public Policy; Scoring; Secondary School Students; *Standardized Tests; Test Interpretation; *Test Results; *Users (Information)

ABSTRACT

Public policy and measurement community issues related to the use of standardized achievement tests for students in kindergarten through grade 12 are discussed, focusing on students and parents as the end-users of test data. The viewpoints expressed are based on 18 years of experience with test score interpretation. It is suggested that technical policy decisions be made to: follow the same groups of tested students over time, rather than simply comparing one group with other similar groups of different years; use a scoring system that allows lay persons to see real measurement of their children's tests and progress over time; and develop, use, and deliver to students and their parents a quasi-official form explaining test results and their significance. In addition, an extensive effort needs to be made to provide for the breakout of scores of Asian and Pacific American students, and the rights of test-takers as consumers should be addressed via standards. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED301599

K-12 Standardized Achievement and Aptitude Testing: the End Users

by John Lum

A testimony written for presentation at the hearing of the Commission on Testing and Public Policy, Saturday, April 11, 1987, at the National Association of Asian and Pacific American Education (NAAPAE) 9th Annual National Conference, Honolulu.

My testimony today will be limited to some concerns I have about K-12 grades standardized achievement and aptitude testing. Part of what I will discuss was alluded to by Richard Schutz as far back as 1980, when he made the following observations on book articles that were supposed to be about evaluation and testing policy:

> I think the papers indicate the status of the measurement community in relationship to public policy quite well. There is still a great gap between the two. Our concerns are those of the measurement and methodology community rather than those of the public and of our elected representatives. The papers say nothing about jobs for unemployed teen-agers, nothing about busing, nothing about education for the handicapped, nothing about tax equalization in financing education....Informed communication between groups as disparate as the educational measurement community and the public policy community is a long way off.*

I want to thank the Commission on Testing and Public Policy for holding this hearing at NAAPAE's (National Association of Asian and Pacific American Education) 9th annual national conference, because it gives one the opportunity to address Schutz' observations. If anything, the gap between the measurement community and the public policy community has widened, if, by the public

TM 012 582

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JOHN B. LUM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

policy community, one includes students and parents. This situation has come about because the measurement field has grown increasingly sophisticated, and students and parents have become increasingly locked out from the measurement technicians.

I view the Commission as being a part of the politics of testing, because it is trying to shape and to influence matters pertaining to testing. The inclusion of the words "Public Policy" in the Commission's title, however, should not lull one into thinking that testing is controlled by written policies and guidelines or influenced by the public. Aside from the technical aspects of test development, much of what happens in testing really stems from practices controlled by the testing elites and gatekeepers; and these practices usually fall into the grey area of informal influencing and shaping, or, in my jargon, non-policy politics.

My viewpoints are based upon 18 years of tinkering with test scores from a variety of perspectives--as a doctoral student, as a large school district evaluator, as an occasional university instructor of educational research and evaluation, as a tutor of students on test awareness, and as a parent of children who have already taken some two dozen standardized tests. Test scores, then, almost take on a life of their own to me.

I think that it is somewhat ironic that after some 13 years since the 1974 edition of Standards for Educational and Psychological

Tests has been published by the American Psychological Association, and after seemingly voluminous extensions of this by works such as the Joint Standards..., 99 and 44/100ths of us in education know almost nothing about the end products of testing--scores and the students who have them. This ignorance about scores and students' what they are, what they mean, what their effects are on individuals, and what can be done about these forms of ignorance--can truly earn itself the title of "the politics of test score ignorance."

One aspect of this ignorance deals with inaccessibility and secrecy. For these, we can point to the professionals. Just considering the panel I tried to pull together yesterday, which attempted to deal with test instrumentation vis-a-vis Asian and Pacific Americans, I and my colleagues made at least one dozen long distance phone calls to get someone from ETS, the Educational Testing Service, to make a presentation on its NAEP (National Assessment of Educational Progress) Program, a major test undertaking. Despite all these calls, we did not receive one return call in three months of efforts. The most help we received was always "try so and so. She or he can help you."

Earlier this school year, when a large urban school district switched versions of the CTBS it was going to use, I asked the regional CTBS sales supervisor if I could purchase a sample packet of the new version or any technical manual that were

listed for sale in the CTBS sales catalog. This supervisor refused to do so, because the school district's research director instructed the sales supervisor not to give anyone in the district any information about this test. Any graduate student could have purchased this test material, but because the CTBS sales supervisor didn't want to lose any business, he didn't tell the research director that the research director's request was not an ethical one.

Going from publishers and developers to research & evaluation agencies, a number of users can relate instances of inaccessibility by major agencies. At last year's NAAPAE Conference, for example, the director of a major center on student testing and evaluation, "solicited" resumes from APA researchers, so that the agency could receive APA input. From my conversations with colleagues who responded and submitted resumes, not one has yet received a positive reply, let alone a cursory or perfunctory acknowledgment of their having sent in their resumes.

The very few examples that I have so far cited, cause public ignorance, because of professional inaccessibility and secrecy. Another cause of public ignorance is, for want of another term, inarticulateness. Here again, one can fault the professional. By this, I mean that most of the reports on standardized testing are simply too technical and sophisticated for any but a few well trained test & measurement experts to understand. To be honest

with you, until three weeks ago, I have never come across an APA test & measurement person, except one in a professional publication. Asians are supposed to be known for their mathematical skills; some even for their statistical skills; and still some even for their statistical skills in educational research & evaluational. But they have not branched off into the subspecialty of educational test & measurement, which not only is a field unto itself, but also, one that is somewhat rapidly growing. Now all this may not be so bad were it not for the fact that most test & measurement people cannot write at a level that even the quasiexpert evaluators can understand. And if the evaluators can't understand, they surely are not able to help the lay person to understand.

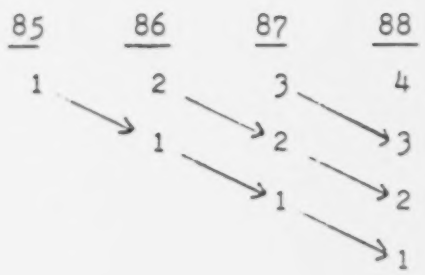
The other side of this coin of high technical skills is that of being quite untrained. If we are at all honest with ourselves, we will find that many school evaluation and testing offices are staffed by people with minimal background in research & evaluation, let alone in tests & measurement. The reports that these people produce make such mortal sins as averaging percentages, utilizing raw score differences, and deliberate obfuscation. Because of their ignorance, some will lean on outside university experts, who themselves do not have the feel of large-scale data gathering or large-scale test administration in a multiethnic situation. The example that comes to mind is a research director who strenuously pushed for the Markov method for predicting

test performance, without accounting for such facts that student movements were not random, and that over time, some score cells would have become empty, i.e., no predictive power at all. This director even strenuously argued that if a student's score was, say, 50, then that score was the probability, thus completely negating the proven concepts of variance and true score analysis, let alone contradict the Markov model that he was advocating.

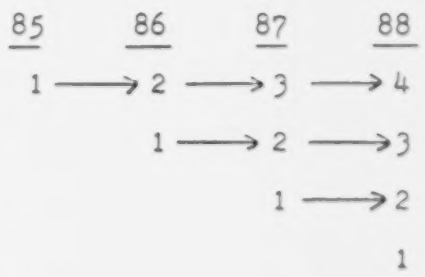
However, I did not come here today simply to say that there are a number of gatekeepers in the game of standardized testing, or that they more often than not fail to serve their constituents for whatever reasons. But, rather, I come to offer a few suggestions of how we can build bridges, even cultural ones (to be in tune with our conference theme) to our real consumers of standardized testing -- our students and their parents. Despite all standards and ethical pontifications different professional groups make about human subjects, very little effort is made on how the human subjects can benefit from the knowledge gained from testing. If human subjects are worried about, the emphasis seems more on how not to hurt them, than on how to help them. Let me try to address this now.

Because of the many difficulties test personnel have in following the same students over time, measurement is often reduced to comparing one against another, instead of one against himself or

herself. The pattern or model that arises is similar to the following:

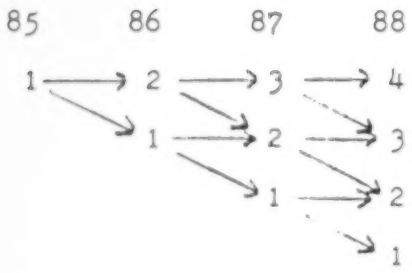


where, for example, one year's first graders are compared to another year's, etc. One knows nothing about how individuals or groups as they progress over the years. And even when there is some attempt to follow the same ones over time, which would look as follows:



there is a need for the kind of scoring that best does this, some form of expanded scale scores, which, over time, covers all the years with just one scale used;for example, the CTBS Tests use a scale of 1-999 (with particular form going from 1-110 for one grade level, and the next form going from 111 - 236 (or something like that), and the SAT Test goes from 200 - 800.

Both of the above kinds of analyses will allow for an evaluator's dream, and would, as a model, look similar to the following:



Related and key to the suggestion that expanded scale scores be used, is that either the school system or the test publisher must develop a simple profile sheet that includes room for K-12 scores. One blank copy, made of high quality paper, since it has to last 13 years, goes to students and their parents. Since test scorers do not have the capacity to follow every test taker over time, at least the students and their parents do have this capacity for themselves. They can stick on their scores over the years. They can show their counselors, teachers, and admissions officers their or their children's individual score trends. They can point out discrepancies between test scores and classroom grades. They can argue for special placement, if warranted. They can, in effect, finally have some control over their scores, which they never did before. These profiles, then, are tools that can raise the level of test score awareness among consumers and educators alike. Schools, of course, should have an official copy of these profiles, too, should the

students' unofficial scores need to be verified. The benefit to schools and researchers is that missing scores become less of a problem. If a centralized researcher cannot or will not give test scores out, which is very often, one can, with some effort, gather individual scores and reconstruct the true picture. In this way, one can even verify how accurate the information is that is cranked out of centralized sources, what with their propensity to introduce selection bias for public relations purposes, i.e., eliminating low scores in their analyses.

I cannot emphasize enough the need for some real action on this relatively simple matter of providing for consumer control of their own scores. Test scores are too delicate a matter to leave solely in the hands of everyone but the ultimate clients themselves. Safety and health regulatory agencies are now legislating and enforcing "right to know" laws. Test professionals should do no less.

To recap this part of the discussion, I suggest that technical policy decisions be made (1) to follow the same groups of people over time, rather than only to compare them with other similar groups of different years, (2) to use a scoring system that allows lay persons to see real measurement of their children's tests and progress over time, and (3) to develop, utilize, and hand over to students and their parents a quasiofficial form that they can possess, and question, and understand sufficiently, and

bargain with, if desired. The implementation of these relatively simple steps could potentially revolutionize the way professionals have treated tests,test scores,and test takers as their own exlusive domain,often at the expense, if not detriment, of the test takers. Tests and test scores are too important to be left only in the hands of the professionals.

Next, I would like to suggest a few more policy and technical related ideas. The first of these is that an all out effort needs to be made to provide for the breakout of Asian and Pacific American (APA) scores, regardless of their numbers. Even small numbers of scores can provide useful information, especially since metaanalytic techniques are now becoming established. The lack of such breakouts has always hampered efforts to help APA students. If such kinds of breakouts are not economically feasible, perhaps a fee can be charged, and/or a tape could be made available, so that the needed information can be extracted. Whatever, test producers and test givers need to make it possible to identify APA test takers.

Next, although I have not yet seen the entire 1985 version of the Standards for Educational and Psychological Testing, I know that there is a new section of ten standards dealing with the rights of test takers. If these ten standards do not yet think of test takers as having consumer rights, and they probably do not, I hope that your project will address this area. Test takers, when

they are thought of at all by test developers, are usually thought of from the perspective of human subjects, not as consumers of a sensitive product. This perspective demands that test developers and publishers think of ways in which test takers can use, if not benefit, from their tests.

Finally, lest we APA educational professionals think that we have no role at all, there are a number of practices, which, while strictly not policy practices, we can do to help our APA students -- item analyses (item difficulty, item discrimination, qualitative item analysis), test wiseness training (how to take tests vs.coaching to the content of tests), use of true score and expected score analyses along with observed score analysis, use of standardized tests other than the one required for practice sake, etc.

In the final examination of what testing is really all about, it's for the student's benefit, not for the school's public relations benefit, nor for the test publisher's, nor for the researcher's. If these others want to benefit, they should do it not at the expense of the students.

The creation of the Commission on Testing and Public Policy comes at a propitious time in the life of our present educational reform movement. Supported by the Ford Foundation, and established by UC Berkeley's Graduate School of Education, this com-

mission will, hopefully, examine testing and its effects from a much wider perspective than has heretofore been done by the reform movement, which seems only to say that better tests need to be developed and that more of these test need to be taken. I look forward to the Commission being an advocate for students and their parents, and to its seriously putting Schutz's caveat into motion:

> Educational measurement and evaluation matters have advanced to a point that they have entered the arena of public policy considerations. This advance brings new responsibilities to all of us in the measurement and evaluation community.**

Reference

* Richard E. Schutz. "Commentary on Policy," Educational Testing and Evaluation: Design, Analysis, and Policy (Eva L. Baker and Edys S. Quellmalz, Eds.), SAGE Publications, Inc. Beverly Hills, 1980, pp.213-214.

** Ibid., p.215.